DOCUMENT RESUME

ED 371 489 EA 025 950

AUTHOR Hilgersom, Karin M.
TITLE Achieving Equity and Excellence through Improved Assessment.
INSTITUTION Oregon School Study Council, Eugene.
REPORT NO ISSN-0095-6994
PUB DATE Apr 94
NOTE 44p.
AVAILABLE FROM Publication Sales, Oregon School Study Council, University of Oregon, 1787 Agate Street, Eugene, OR 97403 ($7, nonmember: $4.50, member: $3 postage and handling on billed orders; quantity discounts).
PUB TYPE Guides - Non-Classroom Use (055) -- Reports - Evaluative/Feasibility (142) -- Collected Works - Serials (022)
JOURNAL CIT OSSC Bulletin; v37 n8 Apr 1994

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Elementary Secondary Education; *Equal Education; *Evaluation Criteria; Evaluation Methods; *Excellence in Education; Foreign Countries; *Standards; *Student Evaluation; Teacher Effectiveness
IDENTIFIERS Australia (Victoria); *Authentic Assessment; Caring; New York City Board of Education; Oregon; *Performance Based Evaluation

ABSTRACT

When evaluating student performance, teachers often find themselves in a quandary. Teachers must judge, sometimes with firm frankness, while also demonstrating caring. Fortunately, assessment in schools is moving away from judgment to embrace the notion of teachers and students sitting beside one another during the teaching, learning, and assessment process. Traditionally, numerically based grades and standardized test scores have been used to set a fixed value on student capability and educator effectiveness. Current debate centers on what comprises effective learning and teaching and how best to measure outcomes. This Bulletin hopes to clarify the recent history surrounding newer assessment forms and to persuade reform-minded educators to consider equity as well as excellence. Chapter 1 briefly examines the relationship between standards and assessment and explores issues surrounding the sophisticated debate on educational assessment. Chapter 2 discusses the difficulties arising as educators balance equity and excellence concerns while designing and implementing assessment tools. Chapter 3 explores various assessment tools being implemented in the United States and Australia. Chapter 4 discusses criteria for evaluating assessment choices (consequences, fairness, transfer and generalizability, cognitive complexity, content quality and coverage, meaningfulness, and cost and efficiency) and makes recommendations. (Includes 44 references.) (MLH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 371 489

Achieving Equity and Excellence Through Improved Assessment

Karin M. Hilgersom

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

✓ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

P. Piele

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Oregon School Study Council
April 1994 • Volume 37, Number 8

BEST COPY AVAILABLE

EA 025 950

OSSC BULLETIN

Achieving Equity and Excellence Through Improved Assessment

Karin M. Hilgersom

Oregon School Study Council
April 1994 • Volume 37, Number 8

ISSN 0095-6694
Nonmember price: $7.00
Member price: $4.50
Quantity Discounts:
10-24 copies - 15%
25-49 copies - 20%
50+ copies - 25%

OSSC STAFF

Philip K. Piele, Executive Secretary
Stuart C. Smith, Editor
Linda S. Lumsden, Associate Editor and Production
Audrey Muller, Publications Sales

OSSC GOVERNING BOARD

Dave Cone, School Board Member, Gresham Grade SD 4
Mike Delvin, School Board Member, North Bend SD 13
James Maxwell, Superintendent, Lane ESD
Philip K. Piele, Professor and Executive Secretary,
Oregon School Study Council
David Conley, Associate Professor, Educational Policy and Management,
University of Oregon
Bill Korach, Superintendent, Lake Oswego SD 7J
Bob Stalick, Superintendent, Greater Albany Public SD 8J
Diane Harr, School Board Member, Oregon School Boards Association

OREGON SCHOOL STUDY COUNCIL

1787 Agate Street
University of Oregon
Eugene, Oregon 97403
(503) 346-5044
Fax: (503) 346-2334

The University of Oregon is an affirmative action, equal opportunity employer.

Preface

The bulk of research and reporting on educational reform revolves around the creation of new educational methods and practices. When we address assessment, however, the discussion is as much about discarding practice as it is about creating practice. Indeed, the word itself can conjure up very different images in educators' minds. Upon hearing about upcoming assessments, one educator might immediately visualize a room full of students, their backs hunched, as they busily bubble-in circles on a scantron. Another may visualize a box large enough for a thirty-page portfolio, a video tape, and a small ceramic vase—all created during a semester by a particular student.

This Bulletin, by Karin Maria Hilgersom, addresses changing assessment practice. It explores a growing resistance to assessments used solely to test and to judge, and highlights efforts to replace such assessment with measurement tools that are authentic, performance based, or both.

Hilgersom recently returned to Liberty Lake, Washington, where she will continue to teach at Spokane Community College. While completing her Ph.D. in Educational Policy and Management at the University of Oregon, spring 1994, she worked as an advisor/instructor for the Educational Opportunity Program (EOP). Hilgersom also worked on the Proficiency-Based Admission Standard Study (PASS), housed at the Oregon State System of Higher Education. She deeply appreciates the support from her colleagues at EOP, and from her colleagues on the PASS project. Hilgersom is also thankful for the support provided by the Center for the Study of Women in Society, and for the opportunity to work with the staff at the Oregon School Study Council.

Contents

	Preface	iii
	Introduction	1
1.	**Promising Standards, Redesigning Assessment**	3
	The Assessment Debate	5
	As for Reformers . . .	8
2.	**Considering both Equity and Excellence in Assessment Redesign**	10
	Neglect of Equity	11
	Standards for Equity: The Educational Opportunity Program	13
3.	**Strategies for Assessment**	15
	Assessment at New York's CPESS Senior Institute	15
	Assessment at Oregon's Cottage Grove High School	17
	The California Student Assessment System	20
	The New Standards Project	21
	Assessment in the Victorian Certificate of Education	22
4.	**Assessing the Assessments**	25
	Challenges	25
	Recommendations	29
	Conclusion	31
	Appendix: CIM Final Assessment (Cottage Grove High School)	32
	Bibliography	35
	Interviews	38

Introduction

Educate—Latin form—educare, to rear

Assess—Latin form—assidere, to sit beside, assist in the office of a judge (*Webster's New Collegiate Dictionary*, ninth edition).

Not surprising to many teachers who relentlessly work to make schools better for children, the term *education* stems from the Latin word *educare*, similar in connotation to the word *care*. I believe that teachers working to reform schools care, and their caring defines their greatest contribution as educators. Caring teachers, and I speak from experience, often find themselves in a quandary. On one hand, teachers must judge, sometimes with firm frankness; on the other hand, they must simultaneously demonstrate caring. Fortunately, assessment in schools is slowly moving away from judgment in the strict sense of the word and starting to embrace what is conveyed by the image of teachers and students "sitting beside" one another during the teaching, learning, and assessment process.

Traditionally, assessment has implied judgment for judgment's sake. Numerically based grades that teachers assign to students and test scores on mass exams sponsored by state education departments and nonprofit national agencies like the College Board are prime examples. Such grades and scores have been used to set a fixed value on student capability and educator effectiveness.

Current debate surrounds the issue of what comprises effective learning and teaching, and how *best* to measure effective learning and teaching. This questioning has led to a national push for excellence in schools, coupled with the reminder that *all* students can learn.

This Bulletin both informs and persuades. I hope to clarify the recent history surrounding newer forms of assessment, along with the philosophical underpinnings of this history. I also hope to persuade educators pursuing school reform—and the changes in assessment practices that substantial reforms usually imply—to consider equity issues as much as, if not more than, excellence issues.

Chapter 1 briefly looks at the relationship between standards and assessment and explores issues surrounding the sophisticated debate on educational assessment. Chapter 2 discusses the difficulties that may arise as educators attempt the dual goals of equity and excellence while designing and implementing assessment tools. Chapter 3 explores various assessment tools being implemented within the United States and Australia, and chapter 4 evaluates assessment choices and practices and offers recommendations.

Chapter 1

Promising Standards, Redesigning Assessment

In a recent interview, Vera Katz, former Oregon state representative and current mayor of Portland, traces newer forms of assessment to educational reform beginning in the 1980s. Stemming from policy of the eighties, recent reforms focus upon setting higher standards for schools. Katz has been involved in such reform at the state and national level for over a decade. Her grasp of the political roots of school reform, and hence the move toward articulating standards and designing new forms of assessment, is both firm and personal. Katz began by sharing her experience as part of the National Center for Education and the Economy:

> I was with a very small group of elite individuals involved in education and the business community. We were involved in writing the report that responded to the *Nation at Risk*—it was called *A Nation Prepared.* Then in 1990, the National Center for Education and the Economy was involved in *America's Choice: High Skills or Low Wages!* with Ira Magaziner and Hilary Clinton, and who's who in the Clinton administration. We reviewed that document and raised all the issues of global competitiveness and asked, "Where are we in the international community?" At that point I realized that if we were just going to tinker with what started in 1987 we would never get there. So we took the entire concept, brought a group together and said, "OK, if you had to blow the system up and put it back together again, what would you do?"

One of the key responses to Katz's question relates to improving standards in education. Standards, or benchmarks, target greater excellence across disciplines. Several national policy-making bodies have begun to specify such standards (National Education Goals Panel 1993, New Standards Project 1993, National Council of Teachers of Mathematics 1991,

National Council for the Social Studies 1993, Geography Education Standards Project 1993, Center for Civic Education 1993, College Board 1983). Many of these policy groups are a response to a host of national reports, including *A Nation at Risk* (National Commission on Excellence in Education 1983), *America's Choice: High Skills or Low Wages!* (Commission on the Sills of the American Work Force 1990), *America 2000* (U.S. Department of Education 1991), *What Work Requires of Schools: A SCANS Report for America 2000* (Secretary's Commission on Achieving Necessary Skills 1991), and Congress's recent passage of *Goals 2000*. Among other claims, these reports suggest that education in the United States has failed to prepare all students for an increasingly complex world of work, or even to provide necessary basic skills. Moreover, these reports imply that American educators should look to countries with stronger economies for better models of teaching and learning.

For Katz, educators have only just begun. She explains:

> So in education you ask yourself, Where are we in relationship to the rest of the world? We know some of that information. We know that we're ahead in some areas and behind in others. In math and science we have a long way to go. When I was in Japan, they have a national curriculum and I took one for the third grade, and I very quickly matched where we were in third grade with where they were in the third grade. It was clear that in some areas we were very on par, and in some areas they were doing fifth- and sixth-grade work in the third grade. So the question for me was, Can we benchmark ourselves to the best in the country, whatever that is, and then to the best in the world?

Creating a new set of standards or benchmarks is relatively simple compared with the task that must follow—designing assessments that accurately measure how well students and educators achieve the standards. Yet even the easier of the two steps—setting the standards—is far from complete. *The Harvard Education Letter* reviews a meeting that occurred in June 1993 of standard-setting leaders representing various disciplines:

> It became clear that there was no agreement on what "standards" meant. Some groups were developing "content" standards to define what students should know, others "curriculum" standards linking teaching activities to the essential core knowledge, and still others "performance" standards focusing on what students should be able to do. (Harvard Graduate School of Education 1993)

Once standards are set, other obstacles remain. Teachers who strive to meet the new standards are often "stymied by outmoded texts and incompatible tests" (Harvard Graduate School of Education n.d.). The success of school reform, assuming the function of school reform is both greater excel-

lence and improved equity, may hinge on the implementation of assessments. For Katz, the answer to greater excellence hinges on the interplay of standards and assessments:

> The New Standards Project is working on the assessments, the euphemism for testing. We [those working in Oregon and on national committees with Katz] always thought that the testing would drive the standard up. If you had the right tests asking the right questions it leads to the teaching methodology to get you there, especially in critical thinking skills or in integration of subject matter. We think that would drive the curriculum. So we're (in Oregon) sort of attacking it at both the assessment level and also at the standards that the CIM is setting. People want to see what the standards really mean in terms of curriculum. I don't know where that is right now. I think the educationalese is getting in the way of that some.

Indeed, the thick "educationalese" surrounding educational reform may bog down educators in their search for better alternatives to assessments. The debate surrounding assessment includes much of the educationalese, as the following section illustrates. Regardless of the jargonized form, however, the content of the debate carries great significance for future assessment. The eventual decisions arising from this debate will certainly affect the opportunity for all students to meet higher standards during the arduous process of school reform.

The Assessment Debate

Authentic assessment and performance-based assessments make use of interpretive modes of assessing learning, often in a "real" context (as opposed to vicarious learning). Examples may include a compilation of a student's best written work in a portfolio, a formal oral presentation, or even a series of mathematical drawings illustrating principles of motion and mechanics.

Critics assert that traditional assessments in the form of standardized tests ensure reliability, whereas newer forms of authentic assessment cannot. Those in favor of such authentic or performance-based assessments typically suggest that validity is strengthened, and that reliability is not essential. Lauren Resnick (1990), among those who favor authentic assessment, asserts that conventional testing methods decontextualize and decompose knowledge, often rendering meaningful assessment unlikely. Grant Wiggins (1993) echoes the claim:

> Today we are failing to negotiate the dilemma. Modern, professionally designed tests intended for national and state use tend to sacrifice validity for reliability. In other words, test-makers generally end up

> being more concerned with the precision of scores than with the intellectual value of the challenge. Thus the forms of testing and scoring used are indirect and generic, designed to minimize the ambiguity of tasks and answers.

An article in a recent issue of *Educational Researcher* devoted almost exculsively to assessment clarifies the crux of the debate. University of Michigan Professor Pamela A. Moss explores "a dialectic between psychometric and hermeneutic approaches." Moss writes:

> Less standardized forms of assessments, such as performance assessments, present serious problems for reliability, in terms of generalizability across readers and tasks as well as across other facets of measurement. These less standardized assessments typically permit students substantial latitude in interpreting, responding to, and perhaps designing tasks; they result in fewer independent responses, each of which is more complex, reflecting integration of multiple skills and knowledge; and they require expert judgment for evaluation. (1994)

Moss further contends that:

> if reliability is put on the table for discussion, if it becomes an option rather than a requirement, then the possibilities for designing assessment and accountability systems that reflect a full range of valued educational goals become greatly expanded. (1994)

Moss believes that a "hermeneutic approach to assessment would involve holistic, integrative interpretations of collected performances that seek to understand the whole in light of its parts, that privilege readers who are most knowledgeable about the context in which the assessment occurs" (1994). In short, students and teachers, as opposed to standardized exams, have the expertise to assess schoolwork in subjective and personal ways. Special projects and portfolios representing a student's unique capabilities may be validly assessed; they permit the ultimate educational goal of improved teaching and learning. Moss asserts:

> I believe the dialogue I have proposed here is not only timely but urgent. We are at a crossroads in education: There is a crisis mentality accompanied by a flurry of activity to design assessment and accountability systems that both document and promote desired educational change. Current conceptions of reliability and validity in educational measurement constrain the kinds of assessment practices that are likely to find favor, and these in turn constrain educational opportunities for teachers and students. A more hermeneutic approach to assessment would lend theoretical support to new directions in assessment and accountability that honor the purposes and lived experiences of students and the professional, collaborative judgements of teachers. (1994)

Samuel Messick, vice president for research for the Educational Testing Service, offers an opposing claim in the same issue. He states:

> Indeed, such basic assessment issues as validity, reliability, comparability and fairness need to be uniformly addressed for all assessments because they are not just measurement principles, they are *social values* that have meaning and force outside of measurement wherever evaluative judgments and decision are made. (1994)

Messick argues that authentic performance assessments are fraught with methodological weakness, namely, construct underrepresentation and construct-irrelevant variance. In short, overreliance on authentic and performance assessment tasks values components of skills to the neglect of complex and well-developed skills. Although Messick obviously favors the accountability and supposed mathematical security of standardized assessment, he grants that "process constructs need to be assessed—not instead of but in addition to complex performances." Messick implies that complex performances are, at this point in time, best assessed by standardized tests. He warns that "it is not just that some aspects of multiple-choice testing may have adverse consequences for teaching and learning, but that some aspects of all testing, even performance testing, may have adverse as well as beneficial education consequences" (1994).

In sum, common ground in the assessment debate has begun to take shape. Assessment may best be a combination of measures, but any measure must be carefully constructed and evaluated. Educators might also ponder the function of any given assessment. Assessments attempting to gauge a school district's effectiveness, therefore aiming to hold schools publicly accountable for their success, are currently designed quite differently fromassessments that aim to improve learning.

In the *Oxford Review of Education,* Willis (1993) mentions that "learning and assessment do not exist in a vacuum." She argues against assessment that neglects this premise:

> If one is interested only in whether students can carry out certain tasks, know certain things or achieve certain objectives, it may be of little concern to know what took place during the learning process itself. What is important is whether they meet objectives rather than why, or why the objectives were not achieved. If, however, one is concerned with improving the quality of learning, and encouraging students to engage in worthwhile activities that stimulate student motivation for future learning it is necessary to look beyond the outcome to examine the process. Rather than assessment being something you *do* to people it is an interactive activity between students and teacher that can play an important role in providing feedback, the aim of which is to improve the quality of future learning.

Willis concludes:

> If students are to be encouraged to engage in high quality learning, assessment must support such learning. To do so, a compatible theory is necessary. It should assume that real learning is active and creative and relevant to real life issues. It is important to develop assessment that reflects this perspective if we want to use assessment to improve learning rather than just measure it.

As for Reformers. . .

Clearly, a national movement that encourages higher quality learning in schools has been under way since the eighties. As for the decade of the nineties, numerous states have adopted laws in which this is the primary aim (See, for example: Colorado's HB 93-1313, 1993; Tennessee's HB 752, 1992; Wisconsin's SB 483, 1991; and Arkansas's Act 236, 1991). In Oregon, the Educational Act for the Twenty-First Century (H.B. 3565, 1991) dictates a wide spectrum of educational change, including the clear articulation of benchmarks and age-specific assessments of those benchmarks. Although many of these assessment tools are standardized and conventional, a move toward performance assessments finds encouragement even from the Oregon Department of Education. Roberta Hutton, an assistant superintendent, explains:

> The whole notion of clear demonstration that skill has been obtained, I think, is an exciting one in terms of kids. In terms of assessment, those in educational circles make a real leap of faith by saying, "If I've taught it, the kids have learned it." And we often have used very poor assessments of that. It's been a very low level of regurgitation that simply isn't going to work anymore.

Policy-makers and educators may interpret the call for higher standards, and thus new assessment options, in two ways. Some educators believe that they must simply change the content of what they teach, so as to improve the chances that their students will score well on objective exams (administered by their state department or by the College Board). Growing numbers of educators, especially those excited about current school reforms, view higher standards as an opportunity to design bold new assessments, usually authentic and/or performance based, that become integral to the learning and teaching processes. Such assessments also include collaborative efforts, where teachers work in teams with other teachers, parents, and community members. The remainder of this Bulletin clearly responds to this latter group of teachers.

For Willamette High School in Eugene, Oregon, such reform efforts

have meant great challenge. Willamette was selected as a pilot school by the Oregon Department of Education to develop a Certificate of Advanced Mastery (CAM) in accordance with the reform act. The principal of Willamette High School, Jim Jamieson, reveals both hope and realism as he ponders changes in assessment:

> The CIM and CAM will be a revolutionary change for students. If it is done right, it will transform their view of the world. It will take them out of the mode, "If I come to class everyday, sit through this, turn in my worksheets, I'm getting a passing grade," which is the way our current educational system is—kindergarten through college. Doing worksheets isn't considered important anymore. Specific skills aren't quite so important anymore. But being able to demonstrate a variety of outcomes through a variety of different performances, work samples, or portfolios—that *would* be important to kids.
>
> Until we get a group here at Willamette who has done it kindergarten through ninth grade, we will have terrific turmoil because we will not be dealing with kids who have enculturated to a new way of doing things. And our problem as teachers is that we haven't done it that way ourselves.

Jamieson's claim seems to be verified by systematic research. Baker and Linn (National Center for Research on Evaluation, Standards, and Student Testing 1994a), summarizing recent case studies on three schools implementing performance assessment, state: "Research indicates that teachers need professional, long-term assistance to implement change. And lots of it."

In short, change is not easy. Teachers may best begin by addressing the functions that their assessments will serve, followed by a time of strategizing that welcomes innovation and empowers those usually closest to students—the students themselves and their teachers.

Chapter 2 begins the trek by offering a conversation, in the hope that educators charged with designing assessment will further the dialogue.

Chapter 2

Considering Both Equity and Excellence in Assessment Redesign

Articles detailing approaches to authentic and performance-based assessment abound. One of the gurus of the movement toward such assessment is Grant Wiggins. On assessing performance, Wiggins states:

> Two key words in this analysis are *context* and *judgment*. Competent performance requires both. It makes no intellectual sense to test for "knowledge" as if mastery were an unvarying response to unambiguous stimuli. That would be like evaluating trial judges only on the basis of their knowledge of law or doctors only on the basis of their recall of biochemistry. What we should be assessing is the student's ability to prepare for and master the various "roles" and situations that competent professionals encounter in their work. (1993)

Wiggins offers great insight on *what* educators should assess students on. He hopes that students learn competent performance through the development of "higher-order habit," which is an "intelligent *proneness*, not a reflex, in an inherently ambiguous situation." To authentically assess "intellectual performance," at least nine factors deserve consideration, according to Wiggins:

1. Engaging and worthy problems or questions of importance, in which students must use knowledge to fashion performances effectively and creatively.
2. Faithful representation of the contexts encountered in a field of study or in the real-life "tests" of adult life.
3. Nonroutine and multistage tasks—*real* problems.
4. Tasks that require the student to produce a *quality* product and/or performance.

5. Transparent or demystified criteria and standards. The test allows for thorough preparation as well as accurate self-assessment and self-adjustment by the student; questions and tasks may be discussed, clarified, and even appropriately modified, through discussion with the assessor and/or one's peers.
6. Interactions between assessor and assessee. Tests ask the student to *justify* answers or choices and often to respond to follow-up or probing questions.
7. Response-contingent challenges in which the effect of both process and product/performance determines the quality of the result.
8. Trained assessor judgment, in reference to clear and appropriate criteria.
9. The search for *patterns* of response in diverse settings. Emphasis is on the consistency of student work—the assessment of *habits* of mind in performance.

Obviously this approach to assessment differs dramatically from standardized testing/assessment methods. Assessment is viewed as integral to the learning process; it may actually engage and empower students. In contrast, the goal of standardized testing is not to teach but to judge—much like a judge in a court of law who has little tolerance for nonconformity of courtroom roles, rules, and procedures. Students rarely feel empowered during standardized tests.

Equity Neglected

The thread running through Wiggins' work and others (Darling-Hammond 1993, Palmer Wolf and others 1992, Resnick 1990, Glaser 1991) primarily addresses the political concern for greater excellence in North American schools. Indeed, they have begun to address excellence issues—quality, competence, high performance, and the like—quite well. What many authors and educators seem to minimize, or completely neglect, is the consideration of issues surrounding equity.

Certainly the issues are complex, and, within the context of everyday life in classrooms, seemingly insurmountable. The adage "all children can learn" implies that schools involved in reform seek and implement ways of helping even the most disadvantaged students to achieve standards of excellence. The phrase may also imply that such disadvantages seem to target certain groups more than others. Ethnic groups, females, and students from poor families may be less academically prepared in certain subjects, or may even be victims of stereotyping and discrimination. For some educators, just saying that "all children can learn" makes it so, or at least the phrase makes

inequity easier to deny. For others, the energy and resources to really accomplish this feat are severely lacking. It is essential for educators who are designing assessments to understand the relationship between equity issues and excellence. In short, true excellence is not achieved if only a select group attains it.

Fortunately, there is enough said about equity issues and how they relate to assessment to begin a necessary dialogue. Performance Assessment Collaboratives for Education (PACE), funded by the Rockefeller Foundation, emphasizes "diversified approaches" to assessments. Additionally, PACE focuses on standards that account for how students' "cultural backgrounds and preparation influence academic goals and performance" (Harvard Graduate School of Education 1993). Moreover, "supports for learning ensure access to resources and opportunities for diverse student populations, to prevent failure, and promote collaboration between schools, families, and community resources on behalf of children."

Serious concerns about equity have been raised by researchers Winfield and Woodard of the National Center for Research on Evaluation, Standards, and Student Testing (CRESST). They argue that elements of President Clinton's recently passed Goals 2000 program may "deepen the already severe educational and economic cleavages that exist in this nation, especially along racial/ethnic lines" (National Center for Research on Evaluation, Standards, and Student Testing 1994b). The authors' research suggests that opportunity for all students to learn will not likely occur without equitable school financing, improved funding for curriculum development, and increased staff development for educators. Winfield and Woodard further claim:

> Only when policy makers consider opportunity-to-learn standards as important as implementing national standards and assessment, will we ensure that those students and individuals historically disenfranchised will share in the American dream of opportunity for educational achievement and economic success. (1994)

Not all state policy-makers neglect equity. Policy-makers at the Oregon Department of Education, for example, are becoming increasingly aware of equity issues, though they must still grapple with how to achieve greater equity with little additional funding. Joyce Reinke, former assistant state superintendent, explains:

> Equity as equal access to programs and equal access for all students regardless of their cultural background or gender certainly has to be one of our primary concerns because we do not have equal access to programs now; we're far from it. We have very resource-poor schools. And we have schools that have a considerable amount of resources and money. Some schools by their location have much

greater access for opportunities than other schools do.

> Many of the equity issues revolve around teacher training and inservicing of teachers, and how they perceive students' capabilities, what students' needs are, what students are capable of doing, and getting away from stereotypes involved. But it's not a short-term project. We now have an advisory council that works with us on developing the Certificates of Initial and Advanced Mastery. The council represents many of the diverse cultural groups, and they give suggestions on how programs should be put together to make sure that we're not excluding or tracking any one subset of any given culture.

Reinke also points out that the greatest difficulty implementing the Oregon school reform act has been an inability to get "information fast enough to everyone and bring everyone on board to get involved in the whole process." Indeed, the process might be so revolutionary for some that any information is difficult to grasp, especially information designed to repair inequity. Roberta Hutton, assistant superintendent of the Oregon Department of Education, considers the adage that "all children can learn":

> If we truly believe that all kids can learn and that we're going to have the same high standards for all kids, then the learning environment has to change and there has to be a broader interpretation of kids' learning needs. The environment has to address those needs with a much broader range of strategies, resources, and materials than we have in the past. The playing field, in essence, must get leveled for kids with those kinds of interventions. I think that piece in and of itself—the whole notion that all kids can learn and that change in the learning strategies and environments facilitates that—puts great pressure on school districts to change. This gets at the very heart of education.

Standards for Equity: The Educational Opportunity Program

Developing standards to ensure equity in schools is as vital as the burgeoning standards used to ensure excellence. Federally funded programs may provide models for how schools can begin to fathom the intensive strategies that might, in Hutton's words, "level the playing field" for many students.

The University of Oregon's Educational Opportunity Program (EOP) provides a good example. Although the program is tailored for adult learners, many features could be applied to learners of all ages. The program works with nearly four hundred students. To qualify for the program, students must meet at least one of the following criteria: be the first in their immediate family to attend college, earn a low income, have a physical

disability, or have a learning disability. Many of the qualifying students meet two of these criteria. As part of the program, students have access to intensive academic advising and individualized assistance, tutoring, workshops, personal counseling, and special courses open only to EOP students; these courses emphasize writing and thinking skills, math strategies, communication skills, and positive self-esteem.

The EOP staff of advisors, instructors, and licensed counselors work as a team of specialists, so students may visit several different staff during any given quarter. This holistic method contrasts with many social agencies that simply divide case files equally among staff. The EOP method encourages collaboration (both written and oral) in a confidential context so that students can be well served.

Obviously, K-12 educators and school counselors cannot easily replicate the University of Oregon's Educational Opportunity Program (EOP) at every school. But there are low-cost lessons on achieving equity that such programs may readily provide.

Chapter 3

Strategies for Assessment

Strategies for authentic assessment and/or performance-based assessment typically encourage the student to compile the culmination of their best work into a subjective format. The use of portfolios is becoming common, though the weight assigned to the portfolio as a significant assessment tool may differ dramatically from school to school or even from class to class.

This chapter provides a smattering of examples of contemporary assessments. Although these examples are far from exhaustive, their intent is to stimulate discussion and help educators envision alternatives. The examples also reveal the potential frustration teachers may feel as they experiment with unfamiliar assessment practices. The examples are drawn from efforts in various states and from efforts in Australia.

Assessment at New York's CPESS Senior Institute

In New York City, at Central Park East Secondary School (CPESS), 450 students engage in an unusual high school completion program. The program reflects the new philosophy of the New York State Curriculum and Assessment Council, a group working to establish "A New Compact for Learning." Assessment shifts from overreliance on standardized testing to comprehensive assessment programs "that include observation of students, evaluation of samples of student work and performance tasks—a major opportunity to motivate more districts to take authentic assessments seriously" (New York State Education Department 1993).

Linda Darling-Hammond (1993) mentions that students at Central Park East Secondary School need not worry about Carnegie Units or the multiple choice Regents examination. Instead, students

> work intensively during one to three years in the CPESS Senior Institute preparing a portfolio of their work that will reveal their

competence and performance in 14 curricular areas, ranging from science and technology to ethics and social issues, from school and community service to mathematics, literature, and history.

This portfolio will be evaluated by a graduation committee composed of teachers from different subjects and grade levels, an outside examiner, and a student peer. The committee members will examine all the entries and hear the students' oral "defense" of their work as they determine when each student is ready to graduate.

The CPESS Senior Institute allows students to design and control the fruits of their labors. Students will likely compile their best efforts, and during the process may study various topics in greater depth. If properly applied, such assessment allows the unique capability and style of every learner to be reflected in the portfolio. By requiring students to defend their

Some Performance

Projects—Projects are comprehensive demonstrations of skills or knowledge. They require a broad range of competencies, are often interdisciplinary in focus, and require student initiative and creativity. Teachers or trained judges score each project against standards of excellence known to all participants ahead of time.

As part of a project, students may be required to conduct a demonstration or give a live performance in class or before other audiences. Projects can take the form of competitions between individual students or groups, or they may be collaborative activities that students work on over time. Science fair projects are a familiar example of this type of performance assessment.

Group projects—Group projects enable a number of students to work together on a complex problem that requires planning, research, internal discussion, and group presentation. This technique is particularly attractive because it facilitates cooperation and reinforces a valued outcome. The California State Department of Education reports success in using group projects.

Interviews/oral presentations—Interviews and oral presentations allow students to verbalize their knowledge. Particularly with younger children, interviews are more likely to elicit informative responses than open-ended, written questions. The 1969 and 1976 National Assessments of Educational Progress (NAEP) Citizenship Assessments used many interview questions.

An obvious example of oral assessment occurs in the foreign languages: fluency can be assessed only by hearing the student speak. As audio and video become increasingly available to record performances, the use of oral presentations for assessment is likely to increase.

Constructed-response questions—Constructed-response questions require students to produce their own answers rather than select from an array of possible answers (as with multiple-choice items). A constructed-response question may have just one correct answer, or it may be more open-ended, allowing a range of responses. The form can also vary, ranging from filling in a blank or writing a short answer, to drawing on a graph or diagram, to writing out all the steps in a geometry proof. Teachers often use constructed-response questions in classroom assessments.

Essays—Essays have long been used to assess a student's understanding of a subject through a written description, analysis, explanation, or summary. Essays can demonstrate how well a student uses facts in context and structures a coherent discussion. Answering essay questions effectively requires critical thinking, analysis, and synthesis.

work before "real" people as opposed to just "school" people, the standard also incorporates presentational skills. The success of CPESS likely hinges upon a highly professional teaching staff who have the ability to coordinate community involvement, and who are aware of students who may be subjected to stereotyping or discrimination—in short, to inequity.

Assessment at Oregon's Cottage Grove High School

Assessment similar to that at CPESS is not limited to urban centers. Located in a rural school district in Oregon, Cottage Grove High School has implemented major reform. Selected as a pilot school to develop Oregon's Certificate of Initial Mastery (CIM), educators at Cottage Grove have made a

Assessment Techniques

Essays and other writing samples may also be used to assess students' composition skills, including spelling, grammar, syntax, and sentence and paragraph structure. Considerable research has been conducted on the standardized and objective scoring of writing assessments. Many states, including Maryland and North Carolina, administer writing assessments at several grade levels.

Experiments—Experiments can be used to test how well a student understands scientific concepts and can carry out scientific processes. Such assessment activities encourage student to "do science" by developing hypotheses, planning and carrying out experiments, writing up findings, using the skills of measurement and estimation, and applying scientific facts and concepts.

A few states are developing standardized scientific tasks or experiments that all students must conduct to demonstrate their scientific understanding and skills. Groups such as the American Association for the Advancement of Science, the National Science Teachers Association, and the U.S. Department of Education's Eisenhower Program are strong advocates for using experiments in classrooms.

Demonstrations—Demonstrations give students opportunities to show their mastery of subject-area content and procedures. Students in a physics class might, for example, demonstrate their understanding of principles of physics in a demonstration using pulleys, gears, and inclined planes. Students in a paramedics course could demonstrate mastery of lifesaving techniques by resuscitating a dummy.

Portfolios—Portfolios are usually files or folders that contain collections of a student's work. They furnish a broad portrait of individual performance, assembled over time. As students put together their portfolios, they must evaluate their own work, a key feature of performance assessment. Portfolios are most common in the subject areas of English and language arts, where drafts, revisions, works in progress, and final papers are typically included to show students' development. A few states and districts use portfolios for science, mathematics, and the arts; others are planning to use them for demonstrations of workplace readiness. Vermont and Michigan are among the states taking the lead on portfolio use for assessment.

Source: Rudner and Boston (1994) (From "Assessing Civics Education," *ERIC Digest Series* (1991) by Lawrence M. Rudner and *Testing in American Schools: Asking the Right Questions* (1992) by the Office of Technology Assessment, Congress of the United States)

variety of changes. Like students in New York City's CPESS program, students in the small town of Cottage Grove will also prepare a senior project on a research topic of their choice, and will present that project to teachers and members of their community.

In the ninth and tenth grades, students at Cottage Grove travel through CIM blocks. A block usually consists of three subjects, and teachers frequently integrate curriculum among the subjects. Since each block lasts several hours, teachers have greater flexibility in the kinds of projects that students can complete, including projects that require computers and sophisticated software. Additionally, students will prepare a CIM project about midway during their high school studies. Students self-select a global issue "with which they feel personal involvement" (Hummel and Parent 1992). Students research the topic, submit a plan and timeline for the project, and perform a self-assessment of this process. Students eventually produce a physical project that they orally describe, defend, and assess.

CIM assessment at Cottage Grove culminates in a Master Portfolio judged by a Board of Review. The board consists of the student, a parent, an advocate, and a CAM strand representative. The Master Portfolio includes twenty-nine outcome measures (See Appendix), most of which are authentic and performance based, but it also includes standardized test scores.

The changes occurring at Cottage Grove have not come easily, and the memories of those involved in the first year of planning vividly depict both highs and lows. Jim Settlemeyer, a CIM block teacher, says that the goal at Cottage Grove was to provide "multiple environments and multiple chances to succeed." Success became defined as going beyond minimum requirements. Settlemeyer recalls:

> Students had a difficult time figuring out what it meant to go beyond minimum requirements. During an evaluation meeting, we decided that instead of just giving guidelines and developing rubrics for minimum requirements for earning a "B," and giving the vague instruction that to earn an "A" you have to go beyond, we thought we would in the first two trimesters, as ninth graders are coming into this program, give them sample rubrics for "A" work. What does it mean to go beyond and really do these open-ended things? Students really didn't know how to do this.

During the year many students began to develop the habits and practices necessary to succeed. Julia Keizur, a counselor at Cottage Grove, notes a shift in her own practice:

> I think we used to say that excellence was having good grades and taking difficult classes. This is hard to do right off the top of your head. I think we're looking more at excellence in process—knowing how to set up your own educational program, knowing how to go beyond the rubric to do more than the bare minimum. I don't know if

> anybody anymore is going to define excellence as knowing how to solve a quadratic equation. Instead, we're concerned with *why* you have to find out how to do that. Students are taking more responsibility for their learning.

Linda Discerni, a language arts teacher deeply involved with the team of teachers who planned the first year of the CIM, recalls the difficulty of assessing students who failed to take that responsibility:

> I don't think anyone has answered the assessment piece of the puzzle. What we've done in simple day-to-day grading is divide activities into key activities and basic activities. Kids have to do all tasks to a level of excellence. But what happens is, they can do it today, or you know, when they are forty years old; it's not like they're going to be penalized as in the old system, such as getting penalized for a late paper. So what we ended up with is a huge bottleneck of students who are probably quite bright but who have done nothing under this new system.

One of the best assessments during the first year, all interviewees agreed, was the "benefit show" produced, directed, and performed by the ninth-grade class. During the third trimester, four groups, each consisting of approximately sixty students, produced a variety show. Students applied for various jobs, including CEOs; department heads of production, business, talent, and marketing; actors; writers; accountants; ushers; and janitors. The students performed for parents and community members and were assessed in part on the basis of the time they spent on the show, and in part by the quality of tasks they completed. Settlemeyer recalls with fondness:

> We stumbled across the key to making this a world-class program where students fulfilled authentic roles. It wasn't just playing *like* you were something; the students *were* actors, stage hands, PR managers. They assumed real roles that tested their ability and made them responsible to their peers.

Three of the four corporate executive officers were female students. Students who played low-status jobs expressed dissatisfaction about their roles, while students in mid- to high-level roles expressed enthusiasm about the benefit show. How the specific roles played by students affected assessment remains unclear and deserves future consideration.

In sum, educators and students at Cottage Grove experienced the trials and tribulations of change, and assessment practice often posed the greatest challenge. A majority of teachers from the first planning team agreed that even with numerous obstacles, time constraints, and moments of great stress, positive changes in teaching, learning, and assessment made the effort worthwhile.

The California Student Assessment System

In California, attempts to dramatically alter state-level assessments are well under way. In addition to "on-demand" assessments (enhanced multiple-choice, open-ended, and essay responses, structured investigations/ experiments), assessments will be "curriculum-embedded" and will rely on portfolios of student work (California County Superintendents Educational Services Association 1993). A newsletter speaking on behalf of assessment reforms states:

> Curriculum-embedded assessment will be high quality, teacher designed tasks that have been thoroughly field tested and made available to teachers statewide, and include writing prompts, group work, investigations, and other methods. Portfolios encompass a wide range of student work and accomplishments collected over the school year, and will be a combination of work developed for a particular class as well as specific pieces of work as part of a statewide portfolio assessment system. (California County Superintendents Educational Services Association).

California's movement toward authentic assessment is intriguing because of its strong stance in favor of new types of assessments. The following principles and beliefs guide efforts of the California Assessment Program (CAP):

1. Curriculum reform will not happen until we fundamentally change our methods and systems of student assessment.
2. Public funding for education calls for public accountability.
3. An integrated assessment system must make good use of technology and provide information that is timely and understandable if it is to be useful and used.
4. The most important single component in the new assessment system will be the statewide performance standards, and the most important outcome of the assessment will be the internalization of those standards in the thinking and work of teachers, students, and parents.
5. One of the most important purposes of the new student assessment system is to nurture local capacity to carry out the more authentic, performance-based assessment at the classroom level—the place where it ultimately matters.
6. The proper role for the state is not to do the actual assessing; that is a local responsibility.
7. Major staff development efforts will be required, and the state must make every effort to provide those efforts to help districts coordinate and maximize the effectiveness of their staff

development resources, and to secure additional flexibility as needed. (California County Superintendents Educational Services Association 1993)

The reformers of school assessment in California hope to decrease the value placed upon state-administered multiple-choice exams, simultaneously increasing the power of teachers and students as the ultimate developers of authentic and performance-based assessment. Yet, like other states, California's policy-makers must operate in a social and political context that demands public accountability. They must continue to clearly communicate successful performance; in essence, they must teach parents and their communities the value of nonexact, and often nonnumerical, assessments of children. Considering an often blind acceptance of easy-to-grasp percentages and two-digit test scores, this poses no simple task.

The New Standards Project

The New Standards Project is a joint collaboration between the Learning Research and Development Center at the University of Pittsburgh and the National Center on Education and the Economy. The center may have been the key force responsible for the national scurry among educational circles to develop long lists of standards at local, state, and national levels. Their prominent report *America's Choice: High Skills or Low Wages!* (National Center for Education and the Economy 1990) set the tone for the educational goal-setting prevalent in this decade. *America's Choice* claims a direct link between a weak educational system and a poor national economy.

Arising from this scenario, the New Standards Project has the primary goal of using "a new system of standards and assessments as the cornerstone of a strategy to greatly improve the performance of all students, particularly those who perform least well now" (New Standards Project 1993). Whether the New Standards Project accomplishes this feat remains largely unclear.

Take, for example, an experimental task labeled "Checkers." Checkers was administered to 1,000 fourth graders in 18 states and 6 urban school districts during spring 1993. Checkers was "one of 70 tasks that were taken by nearly 50,000 fourth and eight graders as part of a pilot examination in math and English" (New Standards Project 1993). After a thirty-minute preassessment activity used to ensure that students understood the purposes and various types of tournaments, students were asked to make a schedule for four children who decide to have a checkers tournament at a school.

Several factors were considered as students planned the schedule: (1) each player wanted to play each of the others at least one time, (2) the tournament was to be completed in one week, (3) the players would play only at lunchtime, (4) the players had two checkers sets that could be used concur-

rently, and (5) one of the players was unable to play on Mondays and Wednesdays. Students were asked to complete the schedule. Then they were asked to create a second schedule that allowed a fifth checker player to join the tournament (New Standards Project 1993).

The students' schedules were assessed by rubrics—criteria for scoring that are designed for use in a "professional collaborative setting: teachers scoring together around a table, with discussions" (New Standards Project 1993). The "Checkers" booklet also provides samples of schedules completed by students and assessed by teachers using the rubric format.

In many ways "Checkers" offers a creative model for assessment. The assessment rubrics, however, "focus on the performance rather than on the performer" (New Standards Project 1993). The assessment, not unlike standardized tests, removes the learner, and the cultural and social context of that learner, from the actual tasks. Although teachers are encouraged to assist students with limited English in understanding the task, little systematic thought on how to deal with such disadvantages is provided, before or during the assessment.

In sum, the New Standards Project may bring positive assessment methods to light. The primary purpose of these kinds of assessment, however, is to better judge—regardless of a learner's readiness for that judgment—and not to better teach.

Assessment in the Victorian Certificate of Education

Schools in Australia are experiencing rapid growth and have "a considerably more heterogeneous student population" (McGaw and others 1990). Australian secondary schools must accommodate a greater variety of learners, and they are responding with reforms that offer a "wider range of curriculum opportunities." Assessment reform in Victoria has culminated in the Victorian Certificate of Education (VCE), a tool that eliminates a "plethora of certificates and relatively uncoordinated diversity of subjects in Victoria."

Vickers (in press) says of the VCE:

> The singular achievement of the VCE is that it has brought about common, state-wide agreements on curricular content, while at the same time allowing considerable local control over both teaching and assessment. It provides a range of options that lead to employment or higher education or both, and its methods of assessment and reporting aim to provide employers and higher education institutions with detailed information, allowing them to make fair and accurate comparisons among students.

Forty-four areas of study compose the VCE. The areas of study include accounting, Australian studies, dance, languages other than English,

legal studies, literature, media, theater studies, mathematics, and physics. Students complete at least four units in each group of studies, and some units occur in sequence.

During studies toward completion of the VCE, students must perform Common Assessment Tasks (CATs). Teachers assess CATs by applying criteria provided by the Victorian Curriculum and Assessment Board (VCAB). When the VCE was first developed, consistency among schools was a goal satisfied by groups of teachers who met to verify work among schools. Vickers explains the purpose of "verification panels":

> During the first two years of the implementation of the VCE, all grade 12 teacher in Victoria were required to attend local Verification Panel meetings, where samples of students work were discussed and VCAB assessment criteria applied. While the formal purpose of Verification Panel meetings was to standardize interpretations of grading criteria, these meetings also served an important professional development objective. Teachers were able to share ideas about the interpretation of study frameworks and CATs and observe the outcomes of other teachers' work. (Vickers in press)

Since that time, Verification Panels have been modified due to political changes and workload concerns. Small schools continue to meet, while larger schools have returned to the standardized achievement test to ensure consistency among schools. Nevertheless, the internal Common Assessment Tasks (CATs) still play a vital role in the overall assessment process.

The VCE administrative handbook details yearly assessments. The 1991 handbook, *CAT 1. Presentation of an Issue*, describes a task where students are required to produce a piece of writing in which they critically analyze, and present a view on, the use of language in Australian media. Students complete the task during the months of May and June; they may consult individuals or reference material while completing the task. According to VCAB (November 1991), the first section

> will be about 700 words in length. It will be an analysis of the use of language (verbal and/or written and/or visual) in the presentation of an issue in no fewer than three and no more than four texts. The texts must have appeared in the Australian media since 31 August of the previous year. At least two of the texts must be from the print media.
>
> A copy of the print media texts used and full bibliographic details of any non-print texts used will be attached to the piece of writing.

The handbook also states that teachers "will monitor the development of the task by sighting plans and drafts of the student's work." Students should be ready to demonstrate understanding of the task and also submit a statement with the completed work declaring "that all unacknowledged work is the student's own." During the 1991 school year, panel verification was applied.

The CAT offers a sophisticated form of educational assessment that can clearly drive up standards. Pursuing this kind of assessment requires collaboration among teachers and high motivation among students. For students who lack that motivation, necessary interventions should be in place so that they too can succeed on assessments as hearty as the one above.

In sum, this chapter has provided examples and critiques of assessment practices that are in the forefront of educational reforms. The next chapter offers additional examples with the focus specifically on recommendations for those who are ready to experiment with newer, and hopefully stronger, forms of assessment.

Chapter 4

Assessing the Assessments

Educators mired in new ideas and bold new practices typically find themselves too busy to pause and reflect upon the effects of their labors. Hopefully, educators undertaking bold new forms of assessment will scheduling time during the hectic year, somewhere and somehow, to systematically evaluate, reguide, and improve their assessment practices. This chapter discusses ways to "assess the assessment." In doing so, it draws upon the experiences of educators.

Challenges

At Cottage Grove High School, Amy Kantrowitz speaks enthusiastically of her new experimental math program. Kantrowitz fought hard for *Interactive Math*; she recalls repeatedly calling its creators at the U of C at Berkeley for permission to test the program at Cottage Grove High School. She explains:

> Interactive Math was written to the standards set by the National Council of Mathematics Teachers back in 1989. Little in the program is given cursory treatment. When students talk about understanding advertisement, we actually do it. When we do communication, we write all the time. We even use stories like *Alice and Wonderland* to learn about shrinking and growing and exponential curves. And there is a unit called Cookies. In Cookies, there are iced cookies and plain cookies. Iced cookies take longer to make but the baker can charge more for them, and make more profit. But those cookies require icing. And they have maximum oven space, and maximum hours they can work in a day. And so the big question for six weeks is how many of each kind of cookie should students hypothetically make to earn the most amount of money. What happens in math like this is that students *really want to know* the answer.

Premium on Teacher's Effort

Although Kantrowitz enjoys the new program, and her dedication to it remains intact after the initial year of experimentation, she notes the incredible amount of time and energy devoted to ongoing, subjective assessment. She confides:

> The program takes me an incredible amount of time. I work all summer long. I work weekends, the snow days I was grading. Each unit has portfolios, and I read up to eleven pages of explanation. The homework does not consist of right and wrong answers so I can't just read out the answer and have them check off the answers. I can't grade the tests on scantron, since I don't use short answers to evaluate work. And if I'm telling the kids it's important that they think on their own, then I have to be willing to understand what they did. Every couple of weeks they have a problem of the week, which is a long problem that is written up. I have to follow their thinking in order to

Criteria for Evaluating

Consequences—How do these performance-based assessments affects the ways teachers teach and students learn? What are the intended and unintended effects of these assessments? For example, teachers who focus primarily on preparing students for an assessment can affect the validity of that assessment (its ability to measure student knowledge). Students who solve a mathematical problem using a memorized algorithm instead of a higher-order thinking skill such as problem solving also can affect the validity.

Fairness—Have fair test items been selected? Do scoring practices reflect students' capabilities fairly? How are we going to use and interpret the results? The shift from standard multiple-choice tests to performance-based assessments raises concern that the performance tasks chosen and the scoring procedures used be appropriate for all students taking the assessment.

Today's students have diverse backgrounds and experiences. Gaps exist between students due to differences in their familiarity with, and exposure to, the test items and in their motivation to perform and learn. Miller-Jones (1989) suggests that teachers use "functionally equivalent tasks specific to the culture and instructional context of the individual being assessed."

To score students fairly, Stiggins (1987) states that it is critical that the scoring procedures used ensure that the "performance ratings reflect the examinee's true capabilities and are not a functions of the perceptions and biases of the persons evaluating the performance." One solution to fairness in scoring is to combine performance-based measurements with multiple-choice questions. However, Linn et al. (1991) believe that "greater reliance on judgmental reviews of performance tasks is inevitable."

Transfer and generalizability—How far do skills in one area transfer to another? What generalizations can we make from the test results? The concern for skill transfer and generalizability is equally important in performance-based assessments and in multiple-choice tests.

Measuring the degree to which skills transfer within a performance-based assessment is heavily dependent upon the task being performed. It is also important to acquire evidence of how students transfer skills to real-world problems.

Cognitive complexity—Does the assessment require students to use higher-order thinking skills to solve and analyze problems instead of memorizing facts and solving well-structured, decontextualized problems?

evaluate it since they are encouraged to think in different ways.

Kantrowitz's experience highlights an incredible challenge for teachers dedicated to meaningful reform. Teachers hoping to use authentic or performance-based assessment must be willing to devote the time and energy to those assessments. Teachers must also value their capabilities and willingly express their criticisms, while guiding the continuous practice and revisions that their students may struggle with.

As Kantrowitz notes, teachers must learn to follow the thinking processes of individual students and how those processes translate into papers, scientific or mathematical projects, or art forms. This personalized view of learners gets at the heart of assessments geared to improving teaching and learning. Unfortunately, many teachers are faced with larger classes as a result of reductions in school staff and may simply not have the same high level of energy that Kantrowitz demonstrates.

Alternative Assessments

Performance-based assessments should focus on developing skills for higher-order thinking, such as problem solving and critical analysis. A student performing a hands-on science problem may not automatically use complex, cognitive processes. Judge the cognitive complexity by analyzing the task. Then, factor in the student's familiarity with the problem and the student's approach to solving it. Does the student's explanation of the process and the results go beyond, "That's how we did it in class"?

Content quality—Is the content of the assessment consistent with the current understanding in the field? Will the content stand the test of time? Most important, is the content worth the student's and the rater's time and effort? To ensure the quality of the content, subject experts may review both the tasks that the student performs and the overall design of the assessment.

Content coverage—Does the assessment adequately cover the subject matter? As Collins, Hawkins, and Frederiksen (1990) note, both students and teachers tend to underemphasize information not covered in the assessment. Also, if the subject matter is not adequately covered, test scores could be misleading, or instructions could be misinterpreted or misunderstood.

Meaningfulness—Does the assessment give students meaningful problems? Do the students gain worthwhile educational experiences? To find out if the assessment is meaningful, analyze the performance tasks and ask students and teachers what they think of them. Finding out how students and teachers perceive and react to the tasks and the assessment provides valuable, systematic information on how meaningful they are.

Cost and efficiency—The standard multiple-choice test is appealing when time and money are limited. Generally, performance-based assessments are more time-consuming and costly, especially for large-scale testing. Can you justify the cost of these more labor-intensive assessments? To keep costs down, the data collection techniques and scoring procedures need to be as efficient as possible.

Source: Rudner and Boston (1994)

Need for Subjectivity

Another dilemma that educators may encounter as they tackle experimental assessments relates to the perceptual boundaries that may be unnecessarily imposed. Grant Wiggins (1993), for example, considers New Jersey's top-score criterion used to assess essays. He states the criterion and then critiques it:

> *Organization/Content:* Samples have an opening and closing. The responses relate to the topic and have a single focus. They are well-developed, complete compositions that are organized and progress logically from beginning to end. A variety of cohesive devices are present, resulting in a fluent response. Many of these writers take compositional risks resulting in highly effective, vivid, responses.
>
> *Sentence Construction:* Samples demonstrate syntactic and and verbal sophistication through an effective variety of sentences and/or rhetorical modes. There will be very few, if any, errors in sentence construction.
>
> *Mechanics & Usage:* Few, if any, errors.
>
> What a bore. Little in this scoring system places a premium on style, imagination, or ability to keep the reader interested. Only the top score description mentions "effective and vivid" responses, instead of those criteria being woven through the whole rubric. Yet we see this limitation in almost every writing assessment, including those of the National Assessment of Educational Progress (NAEP).

Wiggins highlights a problem often exacerbated when policies, once widely disseminated, set perceptual limits. Often teachers forget that they are qualified to make independent choices on how internal assessments should be structured, modified, and eventually scored.

Wiggins reminds us that the beauty of subjective assessments resides in the subjective suggestions and critiques that may be offered to students on various papers and projects. Style, imagination, and ability to keep a reader interested, for example, are qualities that many educators fear they cannot judge. Assessment that moves away from simple judgment and toward improvement allows educators to express their opinions to students on qualities such as style and imagination.

In sum, I believe the greatest challenges facing educators experimenting with authentic-based or performance-based assessment are to find the necessary time and energy and to apply imagination.

Support at Three Levels

The solutions to these challenges will most likely require three levels of support. At the individual level, teachers must support themselves. They

must believe in their professional abilities and expertise, and they must be dedicated to students. Such dedication will also require a willingness to accept staff development ideas as bridges to personal and professional growth.

At the institutional level, educators must work to communicate with parents and communities the necessity for assessment redesign; they must communicate a view of successful learning that personalizes the situation. In essence, communities must be taught that public accountability might best be measured with authentic, ongoing assessment, in addition to standardized, quantitatively scored exams.

Finally, at the societal level, legislators, citizens, and businesses must be willing to increase the tax base in support of schools, if only to keep class sizes small enough so that long-term, meaningful assessment becomes more feasible.

Unfortunately, these are the challenges left to society as a whole with its interlocking parts, challenges that the following recommendations do not begin to address. Instead, the recommendations made below are intended to stimulate thinking during the implementation of educational assessments.

Recommendations

1. *Explore Equity.* Educators working to implement assessments intended to improve learning for *all students* must explore ways of achieving greater equity. When teachers are able to rise above the tendency to deny inequity exists, they can begin to work to improve equity for their students. Such work happens in daily interactions with students and helpful colleagues, and with a struggle to identify one's own biases and perceptual barriers.

2. *Explore Maximum Excellence.* So much has already been said on excellence that there is little more to add. Nevertheless, it is worth emphasizing that educators should question what may already appear to be finished. That is, they should ask, Can assessment rubrics or guidelines, even those handed down from national policy-making agencies, be strengthened? In short, striving for excellence must be an ongoing process.

3. *Guide the Practice.* Educators working on assessment will hopefully guide their efforts thoughtfully and systematically. Mechanisms that demand a reflection on how a specific assessment worked, or did not work, should be in place. Followup is an essential component to guided practice. As Dominick LaRusso, a professor at the University of Oregon, stresses, only perfect practice makes perfect. Educators experimenting with new assessment tools should not be afraid to fail. Ultimately, mistakes can usher in improvements.

4. *Teach Assessment to Parents and Other Citizens.* Assessment introduces a jargon that is unfamiliar to most citizens. Educators must communicate, calling on media sources to help, what authentic and performance-based assessments consist of. Engaging parents and other citizens in the emerging dialogues surrounding reform is a political, but incredibly necessary, task. Fortunately, authentic and/or performance-based assessment can provide enjoyable, colorful, and entertaining examples of what children in schools are capable of accomplishing. Showboating these projects builds the self-esteem of students and educators, and in some cases it increases respect for schools. The four benefit shows held at Cottage Grove that involved the entire ninth-grade class and a significant portion of the small community serve as a fine example of endless possibilities.

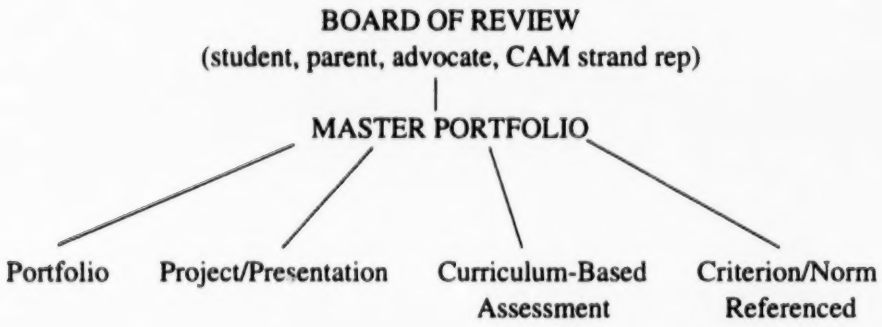
Conclusion

This Bulletin began with a brief glimpse into the history of how educational reforms, and the creation of standards, dovetails with assessment reforms. Next the delicate balance of attempting to achieve both equity and excellence during the design and implementation of assessments was explored. Then a smattering of concrete assessments was offered to inform and to stimulate further planning. Finally, reservations that address the challenges of assessment were shared, and recommendations drawn.

In closing, making imaginative and innovative choices on how to assess students may be at the heart of radically altering and improving schools. Hopefully, through incremental steps and reflective decisions, educators and students will stand firm on their assessment hunches and choices. New forms of assessment that "sit beside," rather than come down from above, may better test the motivational belief that successful learning helps guarantee a successful, happier life.

Appendix

CIM FINAL ASSESSMENT
Cottage Grove High School



• All 29 CIM outcomes are assessed at least once in the assessments listed above.

• The master portfolio will include elements of the four categories—portfolio, project/presentation, presentation, curriculum-based assessment, and criterion/norm referenced tests. This master portfolio will be presented to the board of review when the student and advocate feel the CIM student is ready to exit into the CAM program. Each category of the master portfolio is explained in greater detail in the following pages.

• The assessment process should promote joy/pride and ownership of the portfolio which students, teachers, parents, schools, and districts share.

Portfolio—includes work from three-period integrated block, lifetime fitness block, math, and elective areas.

Project/Presentation—is developed mainly in the three-period integrated block, but will require process and knowledge from other areas, too.

Curriculum-Based Assessment—is a paper and pencil test that covers content in all of the "traditional" areas as well as integral "processes" covered in the CIM program. The test will access more than just "recall" or facts, and should cover a wide range of skills according to Bloom's taxanomy.

Norm and Criterion Referenced Tests—will include such tests as the SAT, OSA COPES, CAPS, and COPS.

Source: *Restructuring for the 21st Century: Implementing the Certificate of Initial Mastery.* Cottage Grove High School, Cottage Grove, Oregon.

CIM PORTFOLIO

PURPOSE: The CIM portfolio is a history and reflection of the student's actions, accomplishments, attitudes, and decisions.

It demonstrates that process, effort, product, and reflection are critical for growth. It focus on improvement of self.

It serves as one component in the assessment of students exiting the Certificate of Initial Mastery program. The other three components include a project/presentation, curriculum-based assessment, and criterion/norm referenced tests.

REQUIRED PIECES IN PORTFOLIO:

- Numbers in parenthesis represent the CIM outcome(s) being addressed.
- All 29 outcomes are addressed except 7, 9, and 22.

1. Letter of portfolio introduction (4, 6, 8)
2. Minimum of two letters of recommendations from adults (25)
3. Essay on ethics, describing student's own ethical code (1, 3, 4, 5, 13, 14, 23, 27, 28, 29)
4. Student's wellness/fitness plan, with a self reflection of the plan (1, 3, 4, 5, 6, 15, 16 , 18)
5. A sample of inquiry and problem-solving activity from across the curriculum with a self reflection of the activity (1, 3, 4, 5, 6, 15, 16, 18)
6. Student's Best piece of writing (4, 6, 8, 12, 13)
7. Reflection of growth with excerpt from reading log (4, 10)
8 Evidence of student's proficiency in the use of technology. (4, 12, 16, 17, 19)
9. Student's best "problem of the week" (4, 11, 15, 17 18, 19, 24)
10. Student's most successful unit portfolio from math (4, 11, 13, 15, 17, 18, 19, 24)
11. Resumé and letter of application to CAM strand (5, 8, 12, 29)
12. Documentation of forty hours civic, community, and/or school involvement (2, 20, 21, 23, 24, 26, 29)
13. Written evaluation of CIM final project
14. Autobiography

OPTIONAL:

1. Videos, photos, slides, or recording of best performance from across the curriculum (6, 9, 12, 23)
2. Videos, Photos slides of student work (6, 7)
3. Special achievements, honors, projects, awards. (6, 22, 24, 25)
4. Any other relevant artifacts as seen fit by student and advocate

•• Each piece of the portfolio has a rubric set of guidelines for the assessment.

Bibliography

California County Superintendents Educational Services Association. "The California Student Assessment System." *California Curriculum News Report* 18, 4 (March 1993): 1-6.

Center for Civic Education. *National Standards for Civics and Government.* Los Angeles: Author, October 1, 1993.

College Board. *Academic Preparation for College: What Students Need To Know and Be Able To Do.* New York: College Entrance Examination Board, 1983.

Commission on the Skills of the American Work Force. *America's Choice: High Skills or Low Wages!* Rochester, NY: National Center on Education and the Economy, 1990.

Cottage Grove High School. *Mathematics Program, 1993-1994.* Unpublished booklet, 1993. 59 pages.

Darling-Hammond, Linda. "Setting Standards for Students: The Case for Authentic Assessment." *NASSP Bulletin* (November 1993): 18-26.

Glaser, Robert. "Expertise and Assessment." In *Testing and Cognition,* edited by Merlin C. Wittrock and Eva L. Baker. Englewood Cliffs, New Jersey: Prentice Hall, 1991.

Geography Education Standards Project. *National Geography Standards.* Indiana, Pennsylvania: National Council for Geographic Education, June 30, 1993.

Harvard Graduate School of Education. "The Push for New Standards Provokes Hope and Fear—and Both Are Justified." *The Harvard Education Letter* IX, 5 (September/October 1993): 1-2.

Harvard Graduate School of Education. "PACE: Performance Assessment Collaborative for Education." Cambridge, Massachusetts: Author. 1-2. No date.

H.B. 3565, Or. Rev. Stat. 1991.

H.B. 752, 97th Gen. Ass'y, 1992 Tenn. Pub. Acts 535.

H.B. 93-1313, 59th Gen. Ass'y, 1st Reg. Sess., 1993 Colo. Legis. Serv. H.B. 93-1313 (West).

Hummel, Howard, and Krista Parent. "Restructuring for the 21st Century: Implementing the Certificate of Initial Mastery." Unpublished booklet, 1992.

Kendall, John, and Marzano, Robert. *The Systematic Identification and Articulation of Content Standards and Benchmarks*." Aurora, Colorado: Mid-Continent Regional Educational Laboratory, January 1994. 269 pages.

McGaw, Barry; Vivian Eyers; Joan Montgomery; Barry Nicholls; and Millicent Poole. *Assessment in the Victorian Certificate of Education*. Victorian Curriculum and Assessment Board, April 1990.

Messick, Samuel. "The Interplay of Evidence and Consequences in the Validation of Performance Assessments." *Educational Researcher* 23, 2 (March 1994): 13-23.

Moss, Pamela A. "Can There Be Validity Without Reliability?" *Educational Researcher* 23, 2 (March 1994): 5-12.

National Center for Education and the Economy. *America's Choice: High Skills or Low Wages!* The Report of the Commission on the Skills of the American Workforce. Rochester, New York, 1990. 209 pages. Abstract. ED 323 297.

National Center for Education and the Economy and the University of Pittsburgh. "Checkers." *New Standards Project*. Washington, DC: Author, August 1993.

National Center for Research on Evaluation, Standards, and Student Testing. "Bumps and Bruises: Challenges to Implementing Performance Assessments in the Classroom." *The CRESST Line* (Spring 1994a): 1-6.

_____. "New Report on Equity and Assessment." *The CRESST Line* (Spring 1994b): 7.

National Commission on Excellence in Education. *A Nation At Risk: The Imperative for Education Reform*. Washington, DC: Author, 1983.

National Council for the Social Studies. *Curriculum Standards for the Social Studies*. Washington, DC: Author, May 3, 1993.

National Council of Teachers of Mathematics. *Professional Standards for Teaching Mathematics*. Reston, Virginia: Author, 1991. 196 pages. ED 344 779.

National Education Goals Panel. *The National Education Goals Report: Building a Nation of Learners*. Volume 1. Washington, DC: Author, 1993. 242 pages.

New Standards Project. "International Benchmarking: Assuring that the New Standards are World Class." Unpublished report, 1993. 26 pages.

New York State Education Department. "Summary of Public Comment on the Curriculum and Assessment Council Interim Report." *A New Compact for Learning*. Albany, New York: Author, 1993.

Palmer Wolf, Dennie, Paul G. LeMahieu, and JoAnne Eresh. "Good Measure: Assessment as a Tool for Educational Reform." *Educational Leadership* 49, 8 (May 1992): 8-13. EJ 444 306.

Resnick, Lauren. "Tests as Standards of Achievement in School." In *The Uses of Standardized Tests in American Education: Proceedings of the 1989 ETS*

Invitational Conference, 63-80. Princeton, New Jersey: Educational Testing Service, 1990.

Rudner, Lawrence M., and Carol Boston. "Performance Assessment." *The ERIC Review* 3, 1 (Winter 1994): 2-12.

Secretary's Commission on Achieving Necessary Skills. *What Work Requires of Schools: A SCANS Report for America 2000.* U.S. Department of Labor. Washington, DC: U.S. Government Printing Office, 1991. 60 pages. ED 332 054.

S.B. 264, 78th Gen. Ass'y, 1991 Ark Acts 236.

S.B. 483, 90th Legis. Sess., 1992 Wis. Legis. Serv. 269 (West).

U.S. Department of Education. *America 2000: An Education Strategy.* Washington, DC: U.S. Government Printing Office, 1991. 52 pages. ED 327 009.

_____. *High Standards for All: Putting Excellence in Education.* Washington DC: U.S. Government Printing Office, 1994.

Vickers, M. "New Directions in the Assessment of High School Achievement: Cross-national Perspectives." In *Linking Schools and Work: Roles for Standards and Assessment,* edited by L. Resnick, D. Jenkins, and J. Wirt. In press.

Victoria Curriclum and Assessment Board. *Assessment in the Victorian Certificate of Education.* Melbourne: Victorian Curriculum and Assessment Board, 1990. 41 pages.

_____. *English Study Design.* Melbourne: Author, November 1991. 44 pages.

_____. *Mathematics Study Design.* Melbourne: Author, September 1990. 70 pages.

_____. *The Victorian Education Certificate (VCE).* Melbourne: Author, 1991.

Willis, Deborah. "Learning and Assessment: Exposing the Inconsistencies of Theory and Practice." *Oxford Review of Education* 19, 3 (1993): 383-402.

Wiggins, Grant. "Assessment: Authenticity, Context, and Validity." *Phi Delta Kappan* (November 1993): 200-08, 210-14. EJ 472 587.

Winfield, Linda F., and Michael D. Woodard. *Assessment, Equity, and Diversity in Reforming America's Schools.* CSE/CRESST technical report #372. Los Angeles: Center for the Study of Evaluation and Center for Research on Evaluation, Standards, and Student Testing, 1994.

Oregon School
Study Council
University of Oregon
1787 Agate Street
Eugene OR 97403

Nonprofit
Organization
US Postage
PAID
Eugene OR
Permit No. 63

OSSC BULLETIN